

P-2213

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-272686

(43)Date of publication of application : 08.10.1999

(51)Int.Cl. G06F 17/30
G06F 17/27

(21)Application number : 10-070688 (71)Applicant : NIPPON TELEGR & TELEPH CORP
<NTT>

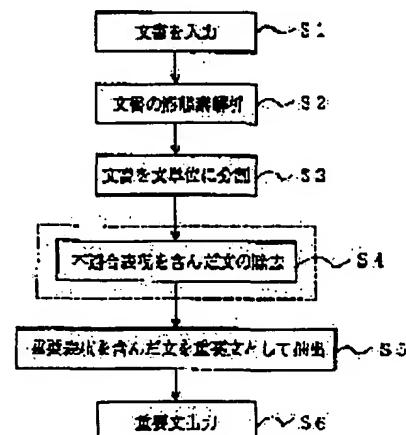
(22)Date of filing : 19.03.1998 (72)Inventor : HORII MUNEYUKI
MATSUOKA KOJI
TAKAGI SHINICHIRO

(54) METHOD AND DEVICE FOR IMPORTANT DOCUMENT SENTENCE EXTRACTION AND RECORD MEDIUM WHERE IMPORTANT DOCUMENT SENTENCE EXTRACTING PROGRAM IS RECORDED

(57)Abstract:

PROBLEM TO BE SOLVED: To easily extract an important sentence from a document with high precision.

SOLUTION: An improper expression table wherein improper expressions are described and an important expression table wherein important expressions are described are prepared; and a morpheme analysis (S2) of an inputted document is taken, the analyzed document is divided (S3) into sentences, and sentences including improper expressions are removed (S4) from the document divided into the sentences by referring to the improper expression table. From the document from which the sentences including the improper expression have been removed, sentences including important sentences are extracted (S5) as important sentence by referring to the important expression table. Here, the process for removing the sentences including the improper expressions is omitted in some cases and the sentences including the important expression may be extracted as important sentence directly from the document divided into the sentences by referring to the important expression table.



LEGAL STATUS

[Date of request for examination] 10.05.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

〔特許請求の範囲〕
〔請求項1〕 文書を入力し、はい込まれた文書を出力する装置を構成するための記憶装置に、
〔特許請求の範囲〕
〔請求項1〕 文書を入力し、はい込まれた文書を出力する装置を構成するための記憶装置に、

された文書を文書単位に分割し(ステップ3)、まず、**要旨文**として不適な表現を記述した不適表現テーブルを段テーブル4-3には、要旨表現が形態素情報、文位置情報とともに記述されている。要旨文抽出部2-4により、重複として記述されている。

参照して、文書位に分離された文書中から不適表現を含んだ文を取り除き（ステップ S 4）、次に、重要表現を記述した重要表現テーブルを参照して、不適表現を含んだ文を取り除いた文書中から重要表現を含んだ文を重要表現テーブル 4-3 に記述されないいずれかの表現を含んだ全文が重要文として文書中から抽出される。抽出された重要文は文書出力装置 3 0に出力される。

から重要な文を抽出する手順を説明する。
【004】文書入口を10で、以下の電子メール文書が入力された場合を考える。

詳しく、文単位に分割された文書中から、重要なテープルを参照して直接重要な表現を含んだ文を原文文として抽出してみよう。

- 2.2. 文書処理装置本体 2.0、文書出力装置 3
- 2.3. 文書処理装置本体 2.0、文書出力装置 3
- 2.4. 文書出力装置 3

- 0、及び、単語検索 41、不適切表現テーブル 42、及び
重要表現テーブル 43を構成する記憶装置 40などで構
成される。
- ・バックアップデータがリストアできません。

この構成は、所謂コンピュータシステムで実現されるものである。ここで、文書入力装置 10 は、キーボード、イメージスキャナ、FD や CD-ROM 等を行いますので、ご出席をお願いいたします。

参考書やDVD等の販売は、実売価格の10%の割引を適用する。また、DVDの販売は、実売価格の10%の割引を適用する。参考書やDVD等の販売は、実売価格の10%の割引を適用する。

る。記憶装置40はハードディスクなどである。
[0010] 図2において、書籍入力装置10は、要
が原文及び部分割点、中段が各章節の箇頭数、下段が各
種語の品詞を表示。書籍入力装置10で入力された文章

〔0116〕文系分析版2.2では、形態解釈された人名は全て形態解釈版2.1にあり、図3と同様に形態解釈される。

参考して形態素解析を行い、この形態素解析された文書を単文分割部 2.2 に送出する。単文分割部 2.2 は、形態素を文書位に分割する。上記入力文 は、以下のように文書位に分割される。ただし、以下では、説明の範

文素解析部2より受け取った文書を文端位に分割する。
ここまでは從来と同様である。英文分割部2.2では、こ
の文書を文端位に分割する。文端位分割部2.3では、文
合上、分割された文の順序に文書を付け替して出す。また、形態素情報は省略する。

2 を参照することにより、單文分割2.2より受け取った文書中から不適表現を込んだ文を取り除く。不適表現

4 システム保全会議のお願いです。

5 社員に納入して頂いた在庫管理システムが今朝ダウ

6. 本状況は以下の通りです。
7. 7月からアクアシャワー4.2切替用面が表示されません。
8. 現在非常モードで作動中です。

〔0012〕重要文書出部24は、重要費現テーブル4
送出される。

11 本日 15時から 17時に三田商事本社会議室で緊
急会議を開く件ないじょうか?

12 3を参照することにより、不運解説部2.3より受取
った文書中から重要表現を含んだ文を抽出する。重要表
現を強調する。

緊急対策会議を行いますので、ご出席をお願いしたいと思います。

8 バックアップデータがリストアできません。
9 年末定期預金で支障があり、早急に対応策を検討しなければなりません。
10 緊急電話を頂けないでどうか?
11 本日15時から17時に三田商事本社会議室で緊急がん金会を行いますので、ご出席をお願いしたいと思います。

1 1.2 申し訳ありませんが、よろしくお願いします。
2 がお届けされる。(「お届けいし」とテーブル1 D 201が
3 マツコ)。
4 00286 したがつて、お届けされる重要な文は、
5 0034システム緊急金庫のお届けいです。

6 1.0 至急電話を頂かないでどうか?
7 1.1 1 日 15 時から 17 時に三田開港大社会課室で緊
8 急對外販売を行いますので、出席をお願いしたいと思
9 います。
10 1.2 申し訳ありませんが、よろしくお願いします。の

述べ、1つでも含まれていた場合、重要として抽出する。

上述の表現のいずれかが含まれているからどうかを調べる。

(0.02.2) (0.5)に、重要表現テーブル4-3の例を図4に示す。重要表現テーブル4-3の構成は、図4の不透表現テーブル4-2の構成と同じである。したがって、重要文曲出部2-4では不透文隊表部2-3と同様の手法で照合を行う。

(0.02.3) 上記文例では、不透文隊表部2-3から受け取った文書のうち、

4. システム緊急会議のお問い合わせです。

の「お問い合わせ」の部分がテーブルD-202にマッチする。

10. 至急電話を頂けないでどうか?

「new」は次の形態群が直後にあることを示す。「new」は次の形態群が直後にあることを示す。「new」は次の形態群が直後にあることを示す。また、「new」は次に形態群がないことを表し、「new」は次に形態群がないことを表す。新規登記、新規登記、品目に沿ける「*」は全てにマッチすることを表す。

1 横須賀相模港 高木津
1.1 本日 15 時から 17 時に三田商事株式会社会議室で緊急会議を行いますので、ご出席をお願いしたいと思
います。

1.2 「横(文末)」の部分が不適表現テーブル 4-2 の
1.3 テーブル 1 D 100 の表現にマッチする。従って、文書
1.4 は、「世話になつた」の部分がテーブル 1 D 102 にマッ
1.5 申し訳ありませんが、よろしくお願いします。

1.6 「が、よろしくお願いし」の部分がテーブル 1 D 1
1.7 0-3 にマッチする。なお、文書号 1-2 は、「、「と「お
1.8 世話になつた」の間に「よろしく」が入っているが、テーブル 1
1.9 0-2 以上により、要文抽出部 2-4 では、上記
1.10 の場合の先頭文は文書号 4。

1.11 「お世話になつた」の部分とマッチするが、文書号 9 は誤植しない。
1.12 とあるので、先頭文でない文書号 9 は誤植しない。

1.13 9 年末時期で業務に支障があり、早急に対応策を検討
1.14 しなければなりません。

1.15 の「を」の部分とマッチするが、文書号 4 は誤植しない。
1.16 「この場合の先頭文は文書号 4」。

1.17 3 文、すなわち、
1.18 4 システム緊急会議のお願いです。

1.19 10 至急電話を頂けないでしょうか?

急封金額を行いますので、ご出席をお願いしたいと思います。
〔0020〕このよーにして、不適文封去部2-3では、
入力カラ書名から文書番号、2、3、1-2の4文が取り除かれて、以下の文書が裏文封去部2-3へ送出される。
4 システム登録会員のお問い合わせです。

5 社員に納入して頂いた在庫管理システムが今朝ダッシュボード、現在非常モードで稼動中です。

6 主な状況は以下の通りです。

7 駐車からアクセスして初期画面が表示されません。

1-2 申し訳ありませんが、よろしくお願いします。
が抽出される。(「お願いし」とテーブル1D201が
マッチ)。

[0026] したがって、抽出される重要な文は、
4-システム緊急会議のお願いです。

1-0 至急電話を頂けないでどうやうか?

1-1 本日15時から17時に三田両本社会議室で緊
急会議を行いますので、ご出席をお願いしたいと思
います。

1-2 申し訳ありませんが、よろしくお願いします。の
4-0となる。

[0027] このように、不適文例部2-3の処理を省
略することと、それほど需要ではない文書号1-2も抽出
されてしまうが、上記文書例では、多少程度が落ちる程
度問題はない。

[0028] 次に、他の具体例として、文書入り会議1
0で、以下の電子メール文書が入力された場合を考え

6 多数の参加、お待ちしております。が抽出される。
[0032] 重要な文抽出部2-4では、図5の重要な文
一テーブル4-3を参照することにより、
2-4月25日に「第3回情報通信省会議」を開催しま
す。
5 参加希望者がございましたら、田中までご連絡くだ
さい。
がそれぞれ「を」(先頭)とテーブル1D204のマ
ッチ、「ご連絡ください」とテーブル1D203のマッ
チで重要な文として抽出される。

[0033] この例で不適文例部2-3の処理を省略し
た場合、文書号2-4では、文書号2が抽出されな
い。これは文書号2が先頭文でなくなるため、テーブル
1D204の文書情報が漏らさないからである。した
がって抽出される重要な文は、
5 参加希望者がございましたら、田中までご連絡くだ
さい。

【0034】以上、本発明の実施例について説明した。ここで、図1に示した本発明の文書要文抽出方法の処理フローは、文書要文抽出プログラムとして、FDもしくはCD-ROM等の記録媒体にコンピュータ読み込み可能な形式で記録し、販売してもよい。この記録媒体に記録された文書要文抽出プログラムを計算機にインストールすると、図2の実施形態での使用が可能になる。

【0035】[発明の効果] 上述のように本発明によれば、重要文書田中です。4月25日に「第3回情報通信学習会」を開催します。今回のテーマは、「エージェント通信」です。佐藤さんとのツーリングには開拓の深いテーマだと思います。参加希望者がございましたら、田中までご連絡ください。多数の参加、お待ちしております。

【0029】形態解析部2.1、単文分割部2.2を経て、
田中です。

す。

3 今回のテーマは、「エージェント通信」です。

4 佐藤さんのグループには関連の深いテーマだと感じます。

5 参加希望者がございましたら、田中までご連絡ください。

6 多数の参加、お待ちしております。

が不適文除去法 2.3 に提出される。

100301 不適文除去法 2.3 では、図 4 の不適表現テ

1. 本発明の文書重要な文抽出技術の実現例の概要

2. 本発明の文書重要な文抽出技術の実現例の概要

3. 四面体構造結果の具体例を示す図である。

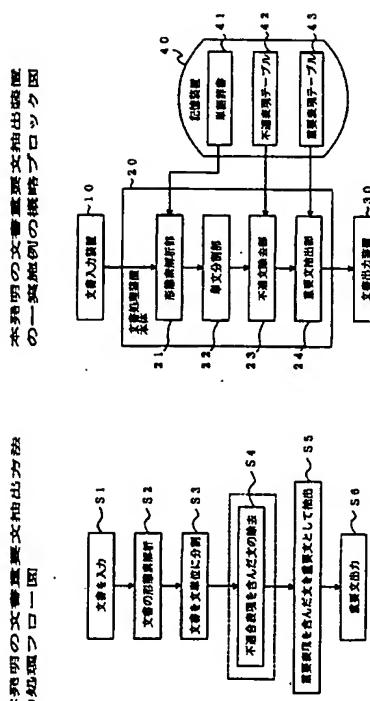
4. 本発明の文書重要な文抽出方法の処理フロー図である。

5. 本発明の文書重要な文抽出方法の処理フロー図である。

6. 本発明の文書重要な文抽出方法の処理フロー図である。

4.0 記憶装置
4.1 單語辞
4.2 不適表現テーブル
4.3 重要表現テーブル

[図 1] 水先明の文文表現文抽出出力部の処理フロー図



三田／桜井／の／竹田／／です／＼
三田／桜井／の／竹田／だ
■水先明(表示) 單語辞 不適表現 重要表現 文書出力

(7)

[図 4]

水先明に用いる不適表現テーブルの内部構成

テーブル ID	属性情報		
	終端記号	終端記号	属性記号
100	+	+	終端記号
101	+	next	終端記号 (12) 終端記号
102	+	next	終端記号
103	+	next	終端記号
104	+	next	終端記号 終端記号
105	+	next	終端記号 終端記号
106	+	next	終端記号 終端記号
107	+	next	終端記号 終端記号
108	+	next	終端記号 終端記号
109	+	next	終端記号 終端記号
110	+	next	終端記号 終端記号

[図 5]

水先明に用いる重要表現テーブルの内部構成

テーブル ID	属性情報		
	終端記号	終端記号	属性記号
200	+	next	終端記号 終端記号
201	+	next	終端記号 終端記号
202	+	next	終端記号 終端記号
203	+	next	終端記号 終端記号
204	+	next	終端記号 終端記号

[図 3]

水先明内部構成の構成